

Betweenness Preference: Quantifying Correlations in the Topological Dynamics of Temporal Networks

René Pfitzner, Ingo Scholtes, Antonios
Garas, Claudio J. Tessone, Frank Schweitzer

Chair of Systems Design, ETH Zurich

Weinbergstrasse 56/58, 8092 Zurich, Switzerland

{rpfitzner, ischoltes, agaras, tessonec, fschweitzer}@ethz.ch

Abstract

Time-evolving interaction patterns studied in different contexts can be well represented by so called *temporal networks*, i.e. networks in which nodes are intermittently connected. In this manuscript we introduce the notion of *betweenness preference* in temporal networks. It captures how likely a certain node is to mediate interactions between particular pairs of its neighboring nodes. We argue that betweenness preference is an important correlation to consider in the analysis of temporal network data. In particular, it allows to assess to what extent paths, existing in time-aggregated, static representations of temporal networks, are actually realizable based on the underlying sequence of interactions. We argue that betweenness preference correlations are present in empirical data sets. We further show that neglecting betweenness preference will lead to significantly wrong statements about dynamic processes on temporal networks.

Many complex systems exhibit dynamically changing interaction topologies. In order to simplify their analysis such systems are typically studied using static, time-aggregated networks. Recent works have argued that properties of dynamic processes evolving on complex networks change significantly when the dynamics of the network topology is taken into account [1]. It has been observed that e.g. spreading processes evolve slower on such temporal networks [2, 3, 4]. As one explanation for this slowdown, the presence of bursty node activity patterns has been suggested [5, 6, 7]. Focusing on *structural properties* of temporal networks, rather than on patterns of node activity, in this manuscript we study to what extent nodes preferentially mediate information flows between particular pairs of neighboring nodes. We propose a measure, *betweenness preference*, to study this preferential mediation. Our study is motivated by the idea that in many networks nodes contact particular other nodes based on the previous contact. One example is the influence of context in information dissemination: work-related emails are more likely to be forwarded to a work-related subset of social contacts. Another example are spatial constraints in disease spreading: an infected individual will not infect all of her (aggregated) social contacts at once. Rather there will be a sequence of contacts, based on the individual's mobility pattern. Here we study the influence of such special classes of dynamical contact patterns arising in temporal networks. We argue that betweenness preference a) is not captured in the time-aggregated

network, b) is present in empirical temporal network data and c) critically influences dynamical processes evolving on temporal networks.

The concept of temporal networks intends to overcome the limitations of time-aggregated network representations. A temporal network is defined as a tuple consisting of a set of nodes $v \in \mathcal{V}$ as well as a set of *events*: $e(v, w, t, l \cdot \Delta t) \in \mathcal{E}$. An event is an interaction between two nodes v and w , starting at time t and with a duration $l \cdot \Delta t$. Here the duration is relative to some smallest unit of discrete time Δt . Based on the time-stamped edges and a discrete notion of time one can construct a *flow-preserving* static representation of temporal networks by *unfolding time* into an additional topological dimension. This construction serves as the basis for our models. We call it a *time-unfolded network*. The time-unfolded networks of two different temporal networks are depicted in the top part of Fig. 1. In the resulting temporal unfolding, we indicate the presence of a possible flow event by an edge $(v_t, w_{t+\Delta t})$, while replacing the original node set \mathcal{V} by a set \mathcal{V}' of *temporal copies* of nodes v_t where $v \in \mathcal{V}$ and $t \in \{0, \Delta t, \dots, L\Delta t\}$ for an observation period of length $L \cdot \Delta t$. Throughout the paper, we assume $\Delta t = 1$ for simplicity. We would like to highlight that similar constructions have been used in the study of temporal networks before [8, 9, 10]. As can be seen in Fig. 1 (top), the two different temporal networks are the same in the time-aggregated representation G_{Agg} (bottom). In G_{Agg} edge weights indicate the number l of discrete time steps in which a particular edge has been active throughout the observation period. In analogy to Statistical Mechanics, one might think of such a time-aggregated network as a macro-state which is compatible with different micro-states, i.e. temporal networks.

Betweenness Preference An important aspect when studying dynamical processes like spreading or synchronization on *static, time-aggregated* networks is that one assumes *transitive paths*. However, this transitivity does not necessarily hold in a (empirical) temporal network that gives rise to the respective time-aggregated network. To illustrate this fact, consider the time-aggregated network G_{Agg} depicted in Fig. 1. If it would represent a static system, information could spread in a transitive way from b via e to g . However, in a temporal network the order in which edges appear imposes an additional constraint: information can only flow along *time-respecting paths* [2]. Hence, in a temporal network underlying G_{Agg} , information can only flow from b to g if there is a time-respecting path, i.e. the connection (b, e) is *followed* by a connection (e, g) . Accordingly, even though the links (b, e) and (e, g) are present in both temporal networks (Fig. 1 top), a time-respecting path between b and g only exists in the left example. In order to capture this transitivity-limiting property of temporal networks in a quantitative way, we study whether certain time-respecting paths are preferentially realized as compared to the time-aggregated perspective. Focusing on the most elementary building block of a time-respecting path, we particularly study what we call a *two-path*, i.e. a path of length two, representing two consecutive edge activations that interconnect three nodes. The statistics of this basic element

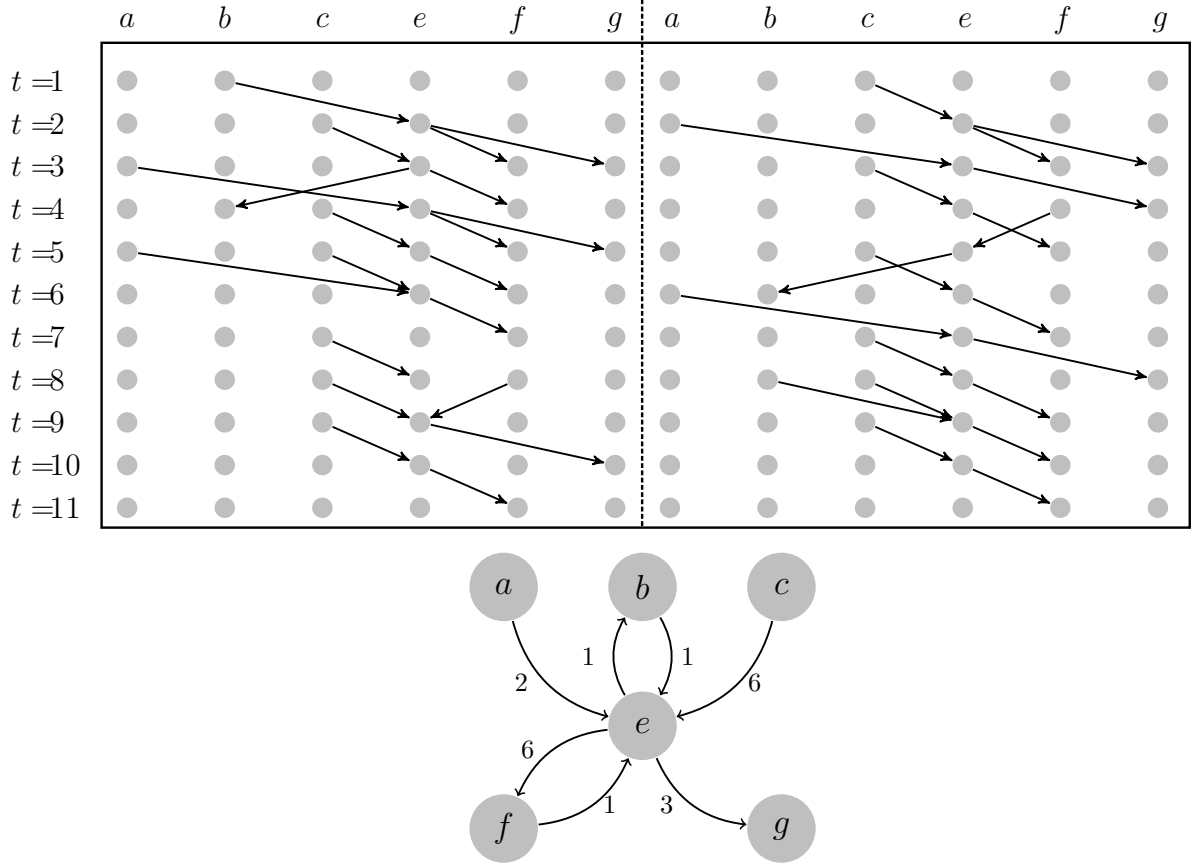


Figure 1: Time-aggregated weighted network G_{Agg} (bottom) and time-unfolded network of two different temporal networks G_{Dyn1} and G_{Dyn2} (top), both of which are consistent with G_{Agg} .

will reveal whether, and to what extent, path-transitivity holds in the temporal network. Based on the introduced time-unfolded representation of temporal networks, we define the elements of a per-node *betweenness preference matrix* $\mathbf{B}^v(t)$ as follows ¹:

$$B_{sd}^v(t) := \begin{cases} 1, & \text{if } (s_{t-1}, v_t) \in E \text{ and } (v_t, d_{t+1}) \in E \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Each matrix element $B_{sd}^v(t)$ captures whether node v_t in a time-unfolded temporal network was in between a source s_{t-1} and a destination d_{t+1} on a two-path $(s_{t-1}, v_t) - (v_t, d_{t+1})$. Based on

¹This definition builds on a notion of time-respecting paths that are comprised of edge activations following each other *immediately*. However, this restriction can be relaxed easily by introducing *self-edges* (v_t, v_{t+1}) for all nodes v and times t .

this, we define the elements of a *time-aggregated betweenness preference matrix* \mathbf{B}^v :

$$B_{sd}^v := \sum_t B_{sd}^v(t) \cdot \left[\sum_{s'd'} B_{s'd'}^v(t) \right]^{-1} \quad (2)$$

and a *normalized betweenness preference matrix* \mathbf{P}^v :

$$P_{sd}^v := B_{sd}^v \cdot \left[\sum_{s'd'} B_{s'd'}^v \right]^{-1}. \quad (3)$$

Essentially, P_{sd}^v is the probability distribution of the two-paths $(s_{t-1}, v_t) - (v_t, d_{t+1})$ over all t . We use this to quantify to what extent v exhibits a preference to interconnect particular pairs of source and target nodes. Based on the concept of mutual information, we define a new temporal-network measure *betweenness preference* as

$$I^v(S; D) := \sum_{\substack{d \in D \\ s \in S}} P_{sd}^v \log \left(\frac{P_{sd}^v}{P^v(s)P^v(d)} \right), \quad (4)$$

where $P^v(s) = \sum_d P_{sd}^v$ and $P^v(d) = \sum_s P_{sd}^v$. In general, $I^v(S; D)$ captures to what extent the knowledge of the source s of a temporal path through v determines the next step d . Or, within a context of information flow, it measures how selective v is in mediating information preferentially between certain pairs of nodes s and d . We note that betweenness preference, as a mutual information measure, is minimal if the random variables S and D are independent. This allows us to calculate the matrix elements P_{sd}^v resulting in $I^v(S; D) = 0$ solely based on the underlying static, time-aggregated network with edge weights w_{ij} :

$$\hat{P}^v(s, d) := p^v(s) \cdot p^v(d). \quad (5)$$

where $p^v(s) = w_{sv} [\sum_s w_{sv}]^{-1}$ and $p^v(d) = w_{vd} [\sum_d w_{vd}]^{-1}$. For the time-aggregated network given in Fig. 1, the joint probability for S and D that results in $I^e(S; D) = 0$ is given as

$$\hat{\mathbf{P}}^e = \left(\begin{array}{cccc|c} a & b & c & f & f \\ 0.12 & 0.06 & 0.36 & 0.06 & f \\ 0.06 & 0.03 & 0.18 & 0.03 & g \\ 0.02 & 0.01 & 0.06 & 0.01 & b \end{array} \right)$$

We will now introduce a simple configuration model to generate temporal networks that are members of a temporal network ensemble defined by a given normalized betweenness preference matrix. Here we will limit ourselves to a subset of possible realizations, in which there is always one edge active per time step and there are only paths of length two realized (as an example consider top left in Fig. 2). The model creates a temporal network from a given \mathbf{P}^v as follows:

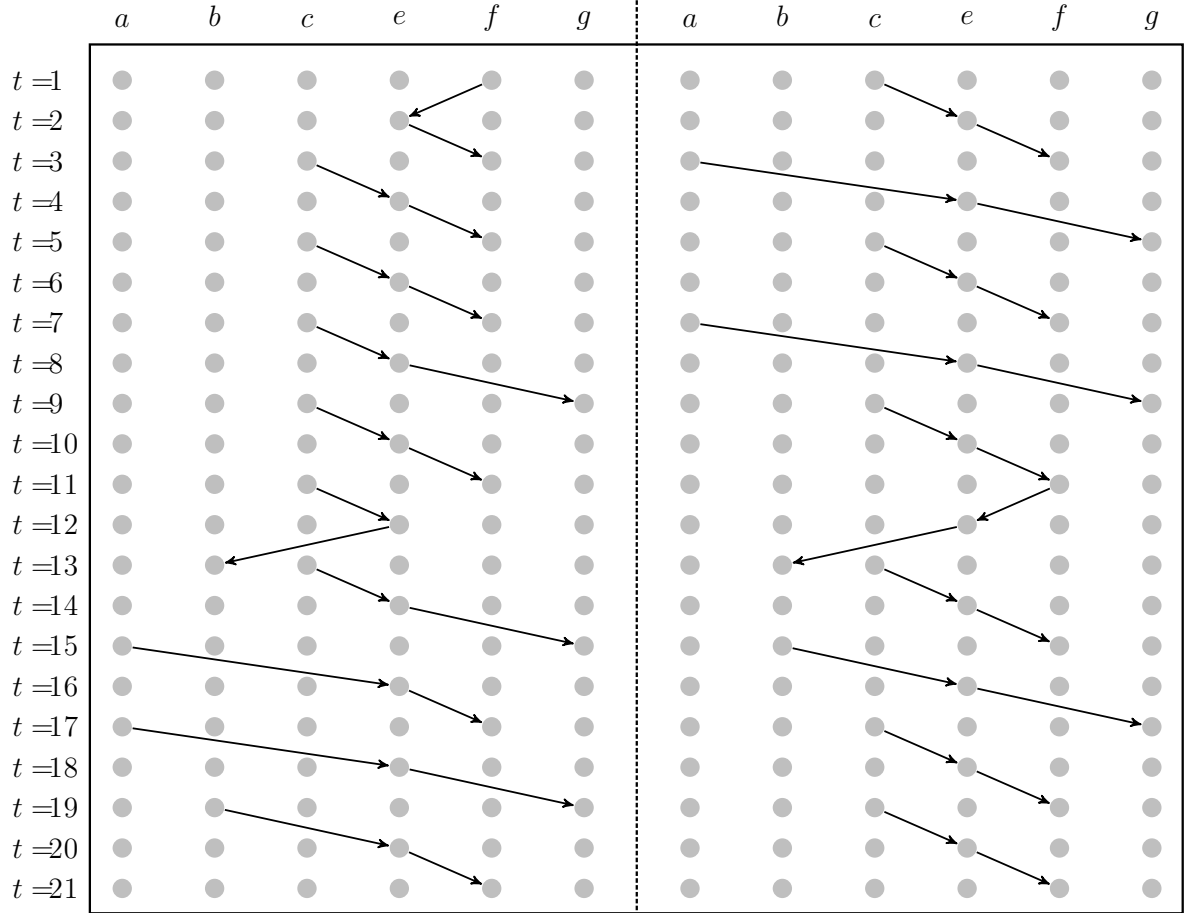


Figure 2: Time-unfolded networks of two temporal networks. Both are consistent with G_{Agg} (Fig. 1). Left: Unfolding of randomly generated microstate; $I^e(S; D) = 0.219$. Right: Unfolding of microstate with high betweenness preference; $I^e(S; D) = 1.295$.

First, we define the number of two-paths N_2 to be realized (which is equivalent to defining the simulation time T). Second, we draw a random two-path $(s, v) - (v, d)$ according to $p(s, v, d) = P_{sd}^v / \sum_v P_{sd}^v$. Third, we create temporal network edges $s(t) \rightarrow v(t+1)$ and $v(t+1) \rightarrow d(t+2)$. We increment $t = t+1$ and $n_2 = n_2 + 1$ (the number of realized two-paths), go to the second step and iterate until $n_2 = N_2$. Having *empirical* temporal network data available, we use this model to create microstates that (i) preserve betweenness preference as well as the macrostate, (ii) destroy other correlations (such as bursty node activations). We call the so constructed temporal network the *betweenness preference preserving case*. Using the configuration model described above, it is also straight forward to construct microstates with low betweenness preference based

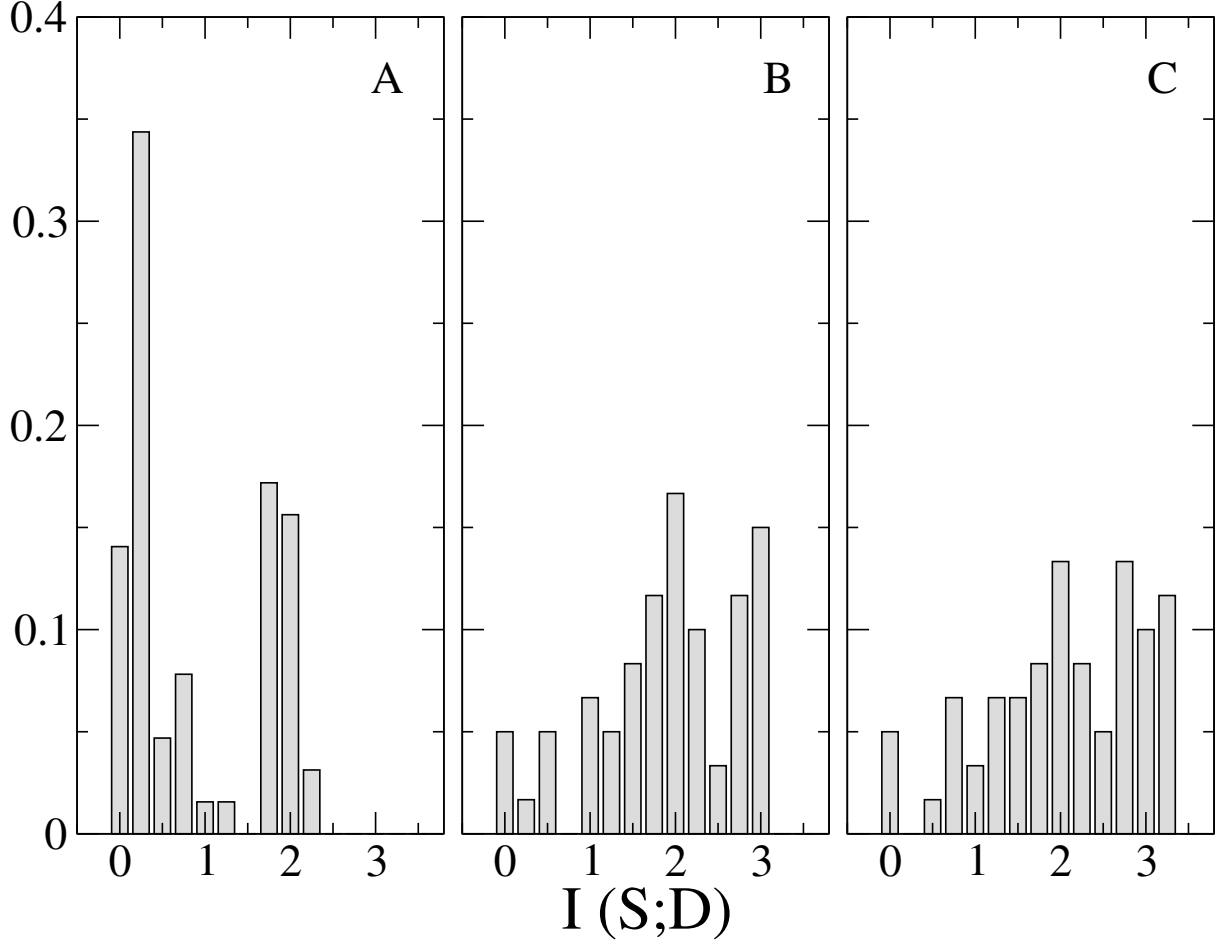


Figure 3: Betweenness preference distribution for (A) a temporal network of the uncorrelated case, (B) the empirical RealityMining data and (C) a \mathbf{P}^v preserving case. All temporal networks have the same length of $N_2 = 50000$ two-paths. (A) and (C) are based on the same empirical data as (B).

on a given macrostate, by using the probability $\hat{P}^v(s, d)$ defined in eq. (5). We call a temporal network created in such a way the *uncorrelated case*, since it only preserves the macrostate, but destroys betweenness preference. To be precise, the uncorrelated case has betweenness preference that is expected from a random microstate of *finite length*. (In a temporal network of infinite length the model would reproduce the limiting case of $I^v(S; D) = 0$.)

Empirical Results As evidence that betweenness preference is an important property in real-world data sets, we show results based on empirical temporal contact networks of the Real-

tyMining Project ². Here we use a one-week subset of the contact network (Sept. 8th to 15th 2004), featuring 64 individuals with 20000 recorded interactions. For details and freely available data see [11]. In Fig. 3 we present the distribution of betweenness preference present in (A) an uncorrelated case, (B) the original data sample and (C) a betweenness preference preserving case. All temporal networks have equal length of $N_2 = 50000$ two-paths. The uncorrelated (A) and correlated (C) case have been created using the empirical data using the configuration model described above. As expected, in the uncorrelated case (A) there is a clear spike around $I(S; T) = 0$, indicating that most nodes show only small betweenness preference. The theoretical expectation of $I(S; T) = 0$ for all nodes in the uncorrelated case is not realized due to finite length of the temporal sequence. Observing betweenness preference in the empirical temporal network (B), one realizes that the distribution is very different from the one in (A). Especially, the distribution in (B) is rather broad, with an average $\langle I(S; T) \rangle = 1.9$ and a median $Q_{0.5}(I(S; T)) = 1.99$, as compared to $\langle I(S; T) \rangle = 0.9$ and $Q_{0.5}(I(S; T)) = 0.5$ in (A). This leads to the conclusion that there is a significant amount of betweenness preference in the empirical contact sequence, that would be missed by assuming the uncorrelated case (A). Hence indicating that there clearly is less path transitivity in empirical temporal network data than would be assumed by the uncorrelated model.

In the third panel (C) we show the betweenness preference distribution of the betweenness preference preserving case. With $\langle I(S; T) \rangle = 2.04$ and median $Q_{0.5}(I(S; T)) = 2.10$, distributions (B) and (C) are very similar. Since we create the model in a statistical fashion based on the normalized betweenness preference matrix \mathbf{P}^v , the two distributions are not completely identical due to finite N_2 . Performing the two-sided Kolmogorov-Smirnov test, we cannot reject the hypothesis that the two distributions are identical with $p = 0.66$. Hence, in this case the model preserves the actual betweenness preference of the real network, whereas all other correlations (e.g. bursty node activities) are destroyed.

Spreading Dynamics in Temporal Networks Turning our attention to the influence of betweenness preference on dynamical processes evolving on temporal networks, we study the dynamics of the paradigmatic SI (Susceptible-Infected) epidemic model. In order to exclusively focus on the impact of betweenness preference correlations, we compare the spreading dynamics on temporal networks created with the uncorrelated model case (A), and the betweenness preference preserving case (C). Additionally to the empirical social contact data, we also consider large synthetic networks that show similar modularity, cluster structure and density as the empirical data [12], and whose nodes have high betweenness preference ³. The time series of the average relative difference between the number of infected individuals $\Delta = (N_u(t) - N_p(t))/(N_u(t))$ in

²The RealityMining project recorded time-stamped social contacts based on proximity sensing technology, with a time-resolution of 5 minutes and over a period of 10 months in 2004/2005.

³For reproducibility, these artificial temporal networks are available at <http://www.sg.ethz.ch/people/prene>

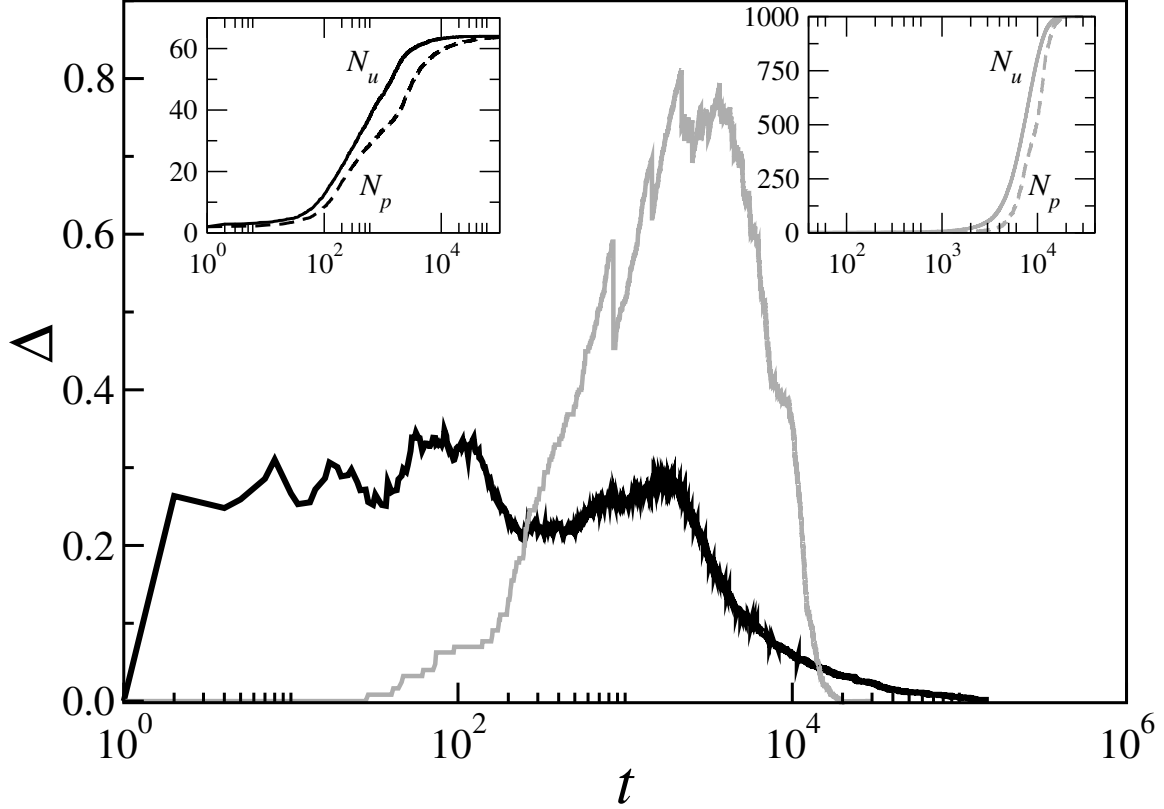


Figure 4: SI-spreading dynamics for temporal networks of empirical RealityMining data (black) and 1000 node synthetic network (gray). Main figure shows time evolution of average relative difference Δ (40 realizations) in SI spreading dynamics between the uncorrelated and the betweenness preference preserving cases. Insets (left: empirical RealityMining data, right: 1000 node synthetic network) show time evolution of number of infected nodes for the uncorrelated (N_u) and betweenness preference preserving case (N_p).

the uncorrelated and the betweenness preference preserving based on the empirical data is shown as the black curve in the main panel of Fig. 4. The solid and dashed curves in the left inset show the evolution of the number of infected individuals for the uncorrelated and the betweenness preference preserving cases respectively. For both cases, equally long temporal networks with 150000 time steps (i.e. 50000 two-paths) have been created and the first active individual was initially infected. The spreading probability was set to $p = 1$. The number of infected individuals clearly follows a typical S-shaped curve (however, notice the semi-logarithmic scaling) in the uncorrelated N_u as well as the betweenness preference preserving N_p case. The slopes in the middle part of the infection dynamics are however clearly different, indicating slower spreading

in the temporal network with non-vanishing betweenness preference N_p . This slow-down is especially indicated by the time-to-saturation which, at least in the cases based on the empirical data (Fig. 4, left inset), is roughly one order-of-magnitude larger for N_p . To substantiate this important effect of large betweenness preference, in the main panel we show the time evolution of the relative differences of infected individuals Δ . Our results clearly show that the uncorrelated model significantly overestimates the average number of infected individuals - at times by up to 35%. Results for the larger synthetic networks (Fig. 4, gray curve and top right inset) confirm these findings and show evidence that in larger systems betweenness preference can have an even more pronounced effects on spreading processes: our simulations indicate that in larger networks with high betweenness preference the slow-down of spreading processes can be as large as $\cong 80\%$. Interpreting Δ as the *error* made when not accounting for betweenness preference, it becomes obvious that taking a time-aggregated perspective of temporal networks, and hence neglecting betweenness preference, can lead to tremendously wrong statements about dynamical processes evolving on networks with dynamic topology. Betweenness preference, as a precise and easy-to-calculate measure, helps to quantify this possible pitfall and to decide whether a time-aggregated network perspective is sufficient. If the nodes of an empirical temporal network show no (or only very small) betweenness preference, realized time-respecting paths are statistically distributed as would be expected from a time-aggregated perspective.

Acknowledgement I. S. acknowledges support by the Swiss National Science Foundation (SNF) through grant CR12I1_125298. C. J. T. and A. G. acknowledge support by the SNF through grant 100014_126865.

References

- [1] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. V. den Broeck, J. Theor. Biol. **271**, 166 (2011),
- [2] P. Holme and J. Saramäki (2011), arXiv:1108.1780v2
- [3] N. Perra, R. Pastor-Satorras, B. Gonçalves, and A. Vespignani, Sci. Rep. **2** (2012),
- [4] N. Perra, A. Baronchelli, D. Mocanu, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, (2012), arXiv:1206.2858v1,
- [5] J. Iribarren and E. Moro, Phys. Rev. Lett. **103**, 038702 (2009),
- [6] M. Karsai, M. Kivelä, R. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki, Phys. Rev. E **83**, 025102 (2011),

- [7] M. Starnini, A. Baronchelli, A. Barrat, and R. Pastor-satorras, Phys. Rev. E **85**, 056115 (2012),
- [8] R. Parshani, M. Dickison, R. Cohen, H. E. Stanley, and S. Havlin, Europhys. Lett. **90**, 38004 (2010),
- [9] H. Kim and R. Anderson, Phys. Rev. E **85**, 026107 (2012),
- [10] B. Mitra, L. Tabourier, and C. Roth, Comp. Net. **56**, 1041 (2012),
- [11] N. Eagle, A. Pentland, and D. Lazer, Proc. Natl. Acad. Sci. USA **106** (**36**), 15274-15278 (2009),
- [12] M. E. J. Newman, M. Girvan, Phys. Rev. E, **69** (**2**), 026113 (2004),